Development of Gene Ontology Tool for Biological Interpretation of Genomic and Proteomic Data

Weimin Feng[1], Geoffrey Wang[2], Barry R. Zeeberg[3], Kejiao Guo[1], Anthony T. Fojo[3], David W. Kane[4], William C. Reinhold[3], Samir Lababidi[3], John N. Weinstein[3], and May D. Wang[1]

[1]The Wallace H. Coulter Biomedical Engineering Department, Georgia Institute of Technology and Emory University, Atlanta, GA 30332-0535

[2]Computer Science and Chemistry Departments, Georgia Institute of Technology, Atlanta, GA 30332

3Genomics & Bioinformatics Group, Laboratory of Molecular Pharmacology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

[4]SRA International, 4300 Fair Lakes CT, Fairfax, VA 22033

Abstract

We have designed and developed a Gene Ontology based navigation tool, GoMiner, which organizes lists of interesting genes from a microarray or a protein array experiment for biological interpretation. It provides quantitative and statistical output files and useful visualization (e.g., a tree-like structure) to map the list of genes to its biological functional categories. It also provides links to other resources such as pubmed, locuslink, and biological molecular interaction map and signaling pathway packages.

Problem and Project Description

Gene expression profiling and other forms of high throughput genomic and proteomic studies are revolutionizing biology and pose new challenges. In addition to the experiment itself and the statistical analysis of results, the biological interpretation is often the most vexing and time-consuming challenge. For example, one generally obtains a list of dozens or hundreds of genes that differ in expression between samples and asks, "What does all of this mean biologically?" The Gene Ontology (GO) Consortium [1] organizes genes into hierarchical categories based on biological process, molecular function, and subcellular localization. In the past, this GO information was queried one gene at a time or batched with a flat format output that does not communicate the richness of GO's hierarchical structure. We have designed and have been developing GoMiner, a bioinformatics tool to automate the functional categorization of gene lists of any length based on the hierarchical structure of the Gene Ontology.

Specifically, GoMiner system was designed to be scalable to full human gene size for biological interpretation of microarray data, with a user-friendly interface. One can input a list of under- expressed, over-expressed, and all genes on the microarray (up to 90,000+). The application features incorporated into the system include (1) graphical representation of hierarchical structure of genes and biological functions; (2) the statistical analysis to indicate enrichment and depletion of biological functional categories with respect to the subset of interesting genes; and (3) linkage to all important biological research resources such as PubMed, LocusLink, molecular interaction map cGAP, biological signaling pathway BioCarta etc. Based on the statistical analysis of enrichment or depletion of biological function categories, one can discover changed gene patterns for biological pathway and regulatory processes study. GoMiner is flexible both because it is coded in Java to be platform-independent, and because it can accommodate either the default GO hierarchy and gene associations or customized versions. The default is the GO Consortium's database of categories and gene associations as implemented on our server. However, the user can, if desired, edit categories and gene memberships using DAG-Edit, the BDGP Gene Ontology Editor Tool (http://www.godatabase.org/dev/editor.html). The edited database can then be accessed by GoMiner from a local server to accommodate domain- and expertise-specific applications. Another important type of flexibility is the wide range of uses. It embraces the full range of genomic and proteomic studies. GoMiner will be downloadable from our website free of charge.

1.   Ashburner M, Ball C, Blake J, Botstein D, Butler H, Cherry J, Davis A, Dolinski K, Dwight S, Eppig J, Harris M, Hill D, Issel-Tarver L, Kasarskis A, Lewis S, Matese J, Richardson J, Ringwald M, Rubin G, Sherlock G: **Gene Ontology: tool for the unification of biology.** *Nature Genet.* 2000, **25**:25-29

Dr. May D. Wang, Biomedical Engineering Department, Georgia Tech and Emory University